GENETIC VARIATIONS in MILK PROTEINS

M. P. THOMPSON

POLYMORPHISM is a term widely used to describe the multiple forms in which either fat crystals or proteins, for example, can exist. Protein polymorphisms (substitutions or deletions of amino acids in the molecule) have been the object of investigation for some years, the study of which has been enhanced by zonal electrophoresis in starch-gel, paper, and other supporting media. Detection of polymorphisms of different proteins by electrophoresis depends in part on the electrical charge carried by the protein. Some noteworthy examples of polymorphisms of functional proteins are the blood proteins, hemoglobin, and transferrin—the genetic distribution of which has been thoroughly investigated. The study of such variations has often been a valuable tool in determining the cause of certain metabolic disturbances.

It is not surprising, then, that the study of genetic variation has been extended to include those proteins which are considered as nontransport, and the usefulness of these studies has become evident. For example, one may distinguish between species of fish by examining the genetic differences in the meat proteins. This is especially useful when a dealer claims that he is selling trout fillets when in fact they are carp fillets. And from the processing viewpoint, there is the possibility that certain genetic forms of protein in food products might impart more desirable characteristics to the product than would other forms.

It is also not surprising that those who have discovered genetic variation in food proteins have regarded the system in which they are discovered as a dynamic, biological system. Often, from a basic or applied research problem on a commodity, discoveries of considerable value to other areas of research, in our case biochemistry, have emerged. Such is the case with milk. Although this remarkable substance is commonly regarded as a nutrient source or something to be made into delicious byproducts, from a chemical point of view it is far more tempting to regard it as a biological fluid. The latter we will do for the remainder of this discussion.

Proteins of the milk system have been broken down into three classes—those contained in the whey, the caseins, and those on the fat-globule surface (membrane). Whey proteins, β-lactoglobulin, α-lactalbumin, and immune globulins have been studied for decades. Until the serendipitous discovery of β-lactoglobulin variation (two forms—*A* and *B*) by Aschaffenburg and Drewry in 1955, no variations of any type had been reported.[1] This astonishing discovery made possible exhaustive physical and chemical studies of the β-lactoglobulins which heretofore had been complicated by a system containing—unbeknown to researchers—two molecular species. Other researchers were prompted to examine the milk system in an attempt to discover further polymorphisms in other milk proteins. Blumberg and Tombs did exactly that when they discovered in Icelandic White Fulani cattle that α-lactalbumin existed in two forms, *A* and *B*, as disclosed by paper electrophoresis.[2] Strikingly, α-lactalbumin is now known to be biologically active; it is the *B* protein of the lactose synthetase system which is responsible for the enzymatic synthesis of lactose in the mammary gland.[3]

Since the precedent of polymorphism had now been established in the whey proteins, Aschaffenburg turned his attention to a study of the possibility of variation in the caseins. Five years after his discovery of β-lactoglobulin variation, he again surprised workers in milk protein chemistry by announcing that β-casein, the second most abundant milk protein, existed in three forms—*A*, *B*, and *C*. Concurrently with the studies of Aschaffenburg,[4] our laboratory was investigating the fraction of casein termed α_s (the fraction that is precipitated by calcium ions, but is stabilized against precipitation by another casein termed χ-casein). The purpose of this investigation was to develop methods of purifying the protein. Eventually, α_s-casein (now termed α_{s1}-) polymorphism was discovered,[5] adding to the growing number of genetic variants discovered in the milk system. Finally, χ-casein, one of the more important and formidable milk proteins, was found to exist in two forms—*A* and *B*.

[1] Aschaffenburg, R., and Drewry, J. Occurrence of Different Beta-Lactoglobulins in Cow's Milk. Nature, 176: 218, 1955.

[2] Blumberg, B. A., and Tombs, M. P. Possible Polymorphism of Bovine α-Lactalbumin. Nature, 181: 683, 1958.

[3] Brodbeck, U., and Edner, K. E. Resolution of a Soluble Lactose Synthetase Into Two Protein Components and Solubilization of Microsomal Lactose Synthetase. Jour. Biol. Chem., 241: 762, 1966.

[4] Aschaffenburg, R. Inherited Casein Variants in Cow's Milk. Nature, 192: 431, 1961.

[5] Thompson, M. P., Kiddy, C. A., Pepper, L., and Zittle, C. A. Variations in the α_s-Casein Fraction of Cow's Milk. Nature, 195: 1001, 1962.

The foregoing briefly sums up the results of 10 years of study on genetic variation. Perhaps now we should ask ourselves what the value of such research has been. From an agricultural viewpoint, we have now been able to seriously consider the chemistry of milk proteins with a consciousness of genetic variants. This knowledge has enabled the chemist to establish how the milk proteins interact to make milk milk, and not a highly disorganized system. From a biochemistry viewpoint, the implications of the studies have been far more significant: studies relating to the mode of inheritance, linkage of genes, genetic code, possible origin of domestic cattle, and differences in chemical-physical properties of protein variants have emerged. Let us consider these.

Mode of Inheritance and Breed Specificity

THE inheritance pattern of the milk proteins—β-lactoglobulin and α-lactalbumin of whey, and α_{s1}-, β-, and χ-caseins—is controlled by a series of multiple-allelic autosomal genes with no dominance: This is to say that if an *A/A* genotype is bred to an all B/B genotype, all of the offspring will be *A/B*. If, however, *A/B* is bred with *A/B* the offspring will be A/A, A/B, B/B.

Extensive investigations on the extent of breed specificity of milk protein polymorphs have presented some interesting results. First, β-lactoglobulin *A* and *B* variation is universal among breeds of cattle studied—Western or domestic breeds (*Bos taurus*) and Eastern (*Bos indicus*), found in India and Africa. However, additional variants have been found in Jersey and French Falamande cattle where β-lactoglobulin *C* and *D*, respectively, have been observed. On the other hand, α-lactalbumin *A* has been observed only in zebu (*Bos indicus*) cattle, while the *B* form is universally prevalent. The importance of this observation will be considered later.

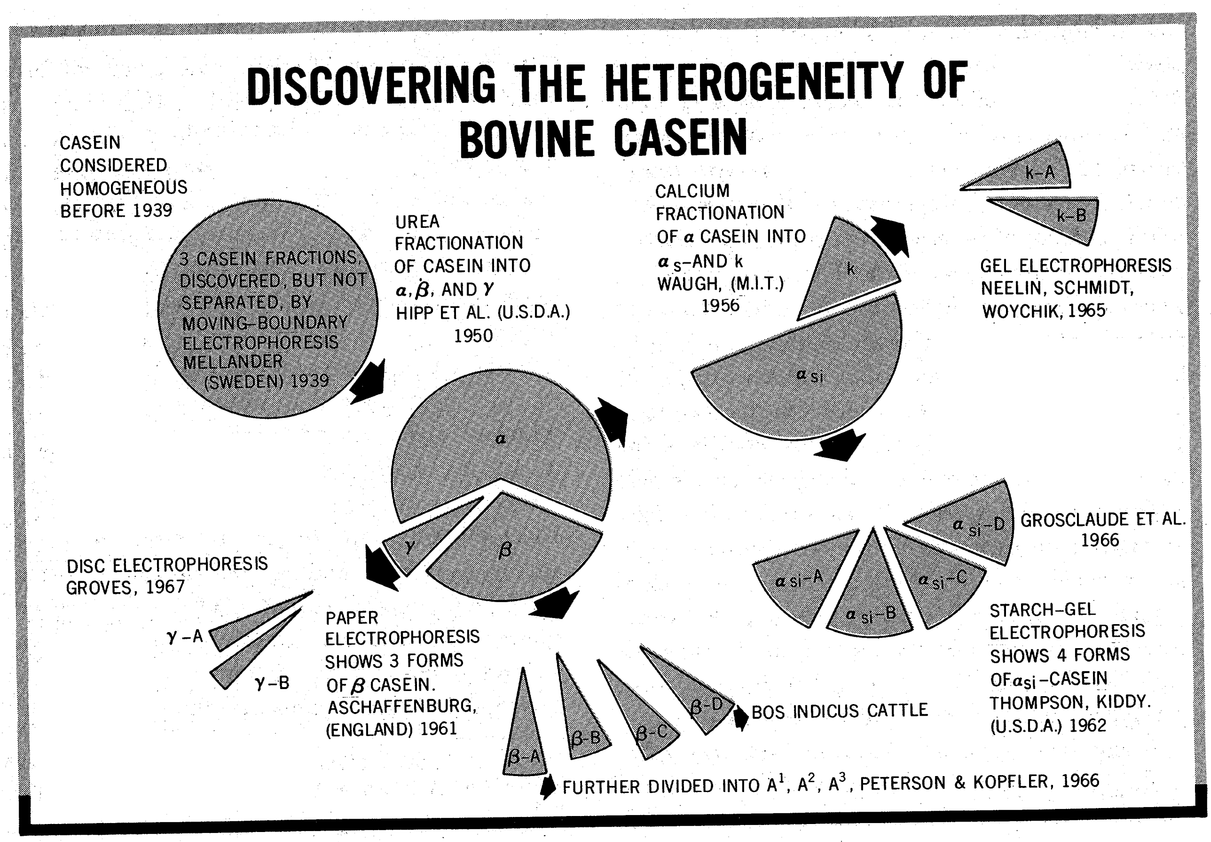
The caseins possess far more breed specific variation than do the whey proteins. β-casein is an excellent example of this. Although the *A* variant (and the author confesses oversimplification of the term "*A* variant," because it is now known to exist in three forms, A^1, A^2, and A^3) is universal among *Bos taurus* and *Bos indicus*, *B* is abundant only in Jersey milks, but is sprinkled throughout Holstein, Brown Swiss, and African and Indian zebu. β-casein *C* is never observed in zebu cattle, but is present in the milks of Guernsey and Brown Swiss, and is totally absent in Jersey, Ayrshire, and Holstein—to name a few. Although zebu are devoid of β-casein *C*, they have an additional variant which has been termed *D*. An interesting point to emphasize here is that, in general, more variation in β-casein is observed in Western breeds of cattle than in the older Eastern breeds.

Zebu cattle predominantly secrete milk containing α_{s1}-*C* caseins, but in European and American breeds the *B* form is present about 90 percent of the time. In all cattle studied to date, with the exception of Ayrshire and Shorthorn, both the *B* and *C* variants are observed; in those breeds we found only the *B* variant. α_{s1}-*A* is the rarest of all the casein variants and, as we will discuss later, it is one of the most unique protein variants found in nature. The *A* protein appears to be limited to a single blood line of Holstein cattle, and perhaps may be a mutant of recent origin. Some Danish workers have reported what appears to be the *A* variant in RDM cattle; however, similar electrophoretic mobility to authentic *A* does not give proof of identity; that is, the variants may differ in amino acid composition.

Lastly, χ-casein variation (*A* and *B*) is universal among all breeds of cattle studied. Interestingly however, the *B* form predominates in Jersey cattle.

Linkage of Genes

LINKAGE of β- and δ-chains of hemoglobin and linkage of egg white proteins has been observed in the occurrence of these proteins. The term "linkage" is defined as the tendency of genes to be passed on, in groups, to the next generation. If, however, phenotypes (or combinations) are seen which do not fit into the exclusion dictated by linkage, then the term "crossing over" is used. With the caseins (and I use the simplest example of this in the milk system), the usual combination of α_{s1}-Cn to β-Cn is α_{s1}-Cn^B, β-Cn^A, although other combinations such as α_{s1}-Cn^C, β-Cn^A are not uncommon particularly in *Bos indicus*. However, English workers in collaboration with the U.S. Department of Agriculture, as well as French researchers, noted that such combinations as



α_{s1}-Cn^{C}, β-Cn^{B} rarely if ever occur. These were regarded as "forbidden" combinations, and their rarity was explained on the basis of linked genes, which simply means that the loci of α_{s1}-Cn and β-Cn are extremely close. The usefulness of linkage information becomes obvious. First, it is another example of the phenomenon; secondly, the information may be used for chromosomal mapping of inherited characteristics of the bovine, a procedure normally carried out in fruit flies (*Drosophila*) and certain bacteria.

The Genetic Code

MOLECULAR biologists and geneticists have attempted to describe, in molecular terms, the factors that allow certain amino acid substitutions to occur within a protein molecule. The language of the cell, the genetic code, is such an attempt. The words of this language are composed of a four-letter alphabet representing the bases of ribonucleic acid (RNA): *A*-adenine, *C*-cytosine, *G*-guanine and *U*-uracil. For example, the amino acid difference between α_{s1}-casein *B* and α_{s1}-casein *C* are +1 glutamic acid and −1 glycine in *B* and −1 glutamic acid and +1 glycine in *C*. This is represented by the triplets GAA/GGA of the genetic code and is a familiar substitution found in hemoglobin, tobacco mosaic virus (TMV) coat protein, and the reverse substitution in tryptophan synthetase. Of course, the *B* and *C* variants of α_{1s}-casein represent an elementary example of the genetic code. The α_{s1}-*A* variant is not describable by the genetic code for it is devoid of at least eight amino acids which are apparently deleted in a single segment of the protein molecule. This is a far more reacting mutation than we have seen with the α_{s1}-*B* and *C* variants, one involving the deletion of a segment of deoxyribonucleic acid (DNA) containing about 24 base units. This deletion ultimately expresses itself in the absence of eight amino acids which we think are in sequence. Consequently, we have a protein deserving of further study because it is not a variant in terms of simple amino acid substitutions or, as will be discussed, does it behave in physical terms like an α_{s1}-casein.

Assignment of triplet coding to β-lactoglobulins *A* and *B*, which differ by two pairs of amino acids, is not easy. However, the *B* and *C* variants of this protein differing by histidine/glutamine, can be described as involving triplets of *CAU/CAA* or *CAC/CAA*. Clearly, however, the milk proteins offer excellent examples of amino acid substitutions which coincide with already observed substitutions and other systems which are predictable in genetic terms.

Origin of Western Cattle

CIVILIZATION of the Western world has been typified by the domestication of cattle and dogs, and in an attempt to decide the route of migration of man from the Middle East, anthropologists have often used the remains of domestic animals and the appearance of artifacts to ascertain that route. Unfortunately, time has a way of eliminating many remains that would be valuable in making such a decision. An alternative method is to examine the living and to observe the frequency of occurrence of a particular substance in an attempt to delineate the origin of a particular plant or animal. This is often done with blood proteins, hemoglobin and transferrin—as well as with blood groups—as a reliable guide for origin. Sen *et al.*,[6] for example, have noted that the frequency of occurrence of hemoglobin *B* in zebu cattle is sufficiently similar to that of the Jersey breed to suggest a possible relationship between Jersey cattle and the more primitive zebu. Furthermore, Stormont[7] remarked that "In my opinion the occurrence of the $A_1D_2Z^1$ blood group in Channel Island cattle (Guernsey and Jersey) is the strongest evidence that one of the ancestral links of those breeds traces to one of the breeds of *Bos indicus*."

Because of the ease of obtaining milk for analysis, we considered that a study of the phenotypes of milk proteins might yet be another valuable way of tracing the origin of Western breeds of cattle. Let us first consider the whey proteins, β-lactoglobulin and α-lactalbumin. The former is found universally among *Bos taurus* and *Bos indicus* in the forms *A* and *B; B* predominates in *Bos indicus*. Because the *C* (Jersey) and *D* (French and German cattle) variants are found in Western breeds only, the value of β-lactoglobulins in delineating the origin of breeds

[6] Sen, A., Roy, Debdutta, Bhattacharya, S., and Deb, N.C. Hemoglobins of Indian Zebu Cattle and the Indian Buffalo. Jour. Anim. Sci., 25: 445, 1966.

[7] Stormont, C., Univ. of Calif., Davis. Personal communication, 1966.

is essentially zero. Although α-lactalbumin *A* is found in zebu cattle, we have not yet observed it in Western breeds. If, however, Jersey and zebu are closely related, as Sen *et al.*, and Stormont suggest, it may yet be discovered in Western breeds. Frankly, I doubt it.

Because we can essentially rule out the value of phenotypes of whey proteins we should, of course, analyze the data obtained from casein phenotyping. Although χ-caseins *A* and *B* are fairly well distributed among *Bos indicus* and *Bos taurus,* a trend clearly emerges from a study of the frequency of occurrence of α_{s1}-caseins *B* and *C* and β-caseins *A, B,* and *D*. The gene frequency of α_{s1}-*C* is high (0.95) in Indian zebu cattle but decreases decidedly to 0.85 in East Africa. However, the *B* variant predominates in *Bos taurus* or the Western breeds. Thus, we see from the subcontinent of India to East Africa to Western Europe a progressive decrease of α_{s1}-*C* and, conversely, an increase of α_{s1}-*B*. Additionally, β-casein *D*—never seen in Western Europe or America—occurs in both Indian and East African zebu. Historically, it has been suggested that African zebu (Boran and Ankole) came from India; our observations strenthen this suggestion.[8]

It is tempting to speculate as to the origin of Western breeds of cattle and their relationship to man over a period of several centuries. But we are admonished to be cautious about our data for two reasons: (1) Our results are too often tainted by preconceived notions, and (2) statistical evidence is rather scarce. More data must be accumulated from a greater variety of breeds of reasonably well-established purity and ancestry. This, I feel, is imperative.

[8] Aschaffenburg, R., Sen, A., and Thompson, M. P., Genetic Variants of Casein in Indian and African Zebu Cattle. Comp. Biochem. Physiol., 25: 177. 1968.

Properties of Genetically Different Milks

As would be expected, from previous knowledge of other polymorphic protein systems, the casein variants are not identical in physical behavior. α_{s1}-Casein *A,* the rarest of the α_{s1}-casein variants, for example, is soluble in 0.4 *M* calcium chloride from 0°–30° C. while the *B* and *C* variants are not. This protein is also stabilized with χ-casein with more difficulty than *B* or *C* over a range of calcium chloride concentrations. Perhaps these observations on isolated α_{s1}-casein become more meaningful when we consider the influence of this variant on the physical behavior of the milks containing the variant. First, cottage cheese prepared from the milk yields a soft curd coagulum which cooks out with difficulty, an obvious economic defect. Second, the α_{s1}-*A* containing milks generally show low stability to heat processing, another obvious economic defect. Third, generally instability of the micellar system can be explained, in part at least, by the low level of water of solvation (1.4 g. water/ g. protein) of casein micelles of α_{s1}-*A* milks as compared to high solvation (1.8–2.2 g. water/ g. protein) for normal milks. Colloidal (micelle) stability depends in part on the level of water solvation as well as on the electrical charge of the colloid.

Evidently the α_{s1}-*A* variant is the most peculiar in chemical composition and physical behavior in the isolated and natural systems. The normal variants do not behave as peculiarly. I would suggest, therefore, that processing variations in individual milks are due in part to protein polymorphisms as well as to the variation of natural salts in milk.

Currently we are engaged in research on the studies of casein micelle stability as affected by genetic variation and quantity of casein components, on the one hand, and on the other, the influence of the above on the binding of calcium and phophorus in the micelle system.